

AMENDMENTS TO THE SPECIFICATION

Please replace the original paragraph [0027] on page 11 with the following amended paragraph:

[0027] In certain applications, seasonal trends and peak periods can be taken into account by “detrending” the sampled data. For instance, if the data stream 12 being observed by the monitoring system 10 comprises disk access rates at a corporation, the access rate may regularly and predictably show an increase at certain times of the day (e.g., 9:00am). Such a change may be considered part of the expected behavior of the system, and indeed, a failure to rise might be considered an event worthy of note. To avoid the detector 22 raising the alarm [[20]] 28 upon seeing this expected change, the data stream 12 may be constructed from the sampled data by computing the difference between the sampled data and data sampled at saliently “the same time” in other periods in the past.

Please replace the original paragraph [0042] on pages 17-18 with the following amended paragraph:

[0042] Referring initially to Fig. 2, the data monitoring system 34 includes a trainer 20 configured to set a threshold 36 that will be used to determine whether the data in the testing window 18 contains a salient change or is saliently different from the data that precedes it in data stream 12. As previously described, to determine a value for a sensitivity parameter 26, such as the threshold 36, the trainer 20 generates a number of sequences 24. As can be appreciated, increasing

the number of data points and generating a number of sequences used to determine the threshold 36 increases the reliability and validity of the threshold 36. The data monitoring system 34 further includes a more specific embodiment of the detector 22, having a scoring function 38. Generally, the scoring function 38 is an algorithm that takes a sequence of data points from either the testing window 18 (during testing) or from sequences 24 generated based on the training window 16 (during training) and computes a score for the sequence. During the training period, the scoring function 38 receives the sequences 24 generated from the data in the training window [[18]] 16 and computes, for each, a score 40. The score 40 may be, for example, the maximum value in the corresponding sequence 24 generated from the data in the training window [[18]] 16, a statistical parameter of the sequence 24, or a more complex value such as the value computed by a Cumulative Sum (CUSUM) algorithm on the sequence 24.

Please replace the original paragraph [0043] on pages 18-19 with the following amended paragraph:

[0043] Referring now to Fig. 3, the generation of the sequences 24 and the selection of the threshold 36 is further illustrated. As previously described, the trainer 20 receives data from the training window 16. The data in the training window 16 is assumed to be unchanging or uninteresting and to have been drawn from some statistical distribution. While this actual distribution from which the data in the training window 16 was drawn is likely to be unknown, in one exemplary embodiment, the trainer 20 may infer a statistical distribution of a known type that appears to model the data in the training window 16, and this inferred statistical distribution may be

used to generate the sequences 24. Accordingly, any sequences 24 generated from the data in the training window 16 can be said to come from the same statistical distribution and therefore also be unchanging or uninteresting. To increase the reliability and validity of the threshold 36 the trainer 20 generates a number of sequences 24. Specifically, the trainer 20 will generate k sequences 24 of length n. In the exemplary embodiment being described, the statistical distribution is assumed to be a discrete distribution containing all of the of the data values actually seen in the training window 16. Further, the distribution includes only data values that are actually present in the training window 16 and includes them at the frequencies in which they appear in the training window. Still further, the elements are assumed to be independently drawn. The sequences 24 are therefore generated by sampling the data from the training window 16, as indicated in blocks 44 and 46 of Fig. 3. It will be apparent that the numbers k and n need not be invariant and different generated sequences 24 may have different lengths.

Please replace the original paragraph [0050] on page 22 with the following amended paragraph:

[0050] Alternatively, rather than sorting the scores and selecting the score with the highest tolerable false positive rate from the sorted scores, a binary search through possible thresholds 36 may be implemented to find the score representing a target false positive rate. Initially, a hypothetical threshold 36 may be selected and sequences 24 may be generated to estimate the false positive rate using this threshold 36. Once two thresholds 36 that bracketed the target rate are determined, a binary search may be performed, repeatedly bisecting the bracketed region and

adjusting it based on which side of the midpoint's false positive rate the target false positive rate is found.